

WE CLAIM:

1. A method for digital data gathering in response to a query, comprising: conducting concurrent searching of structured and unstructured data sources, and preselecting data sources most likely to contain a valid response to the query before submitting the query to the data sources.

2. The method for digital data gathering in response to a query, according to Claim 1, further comprising: combining results from said structured and unstructured data source searches.

3. The method for digital data gathering in response to a query, according to Claim 1, further comprising: combining results from said structured and unstructured data source searches and sorting the results for providing a direct answer.

4. The method for digital data gathering in response to a query, according to Claim 1, further comprising: combining structured data sources into a physical data warehouse with a meta-data repository.

5. A method of digital data gathering for providing an answer to a natural language question, comprising:

- a) accepting input of a natural language question;
- b) identifying the relevant concepts of the natural language question;
- c) assembling the relevant concepts of the natural language question into a query;
- d) identifying a structured data source likely to contain an answer to the query;
- e) performing a first search of the query in the structured data source;
- f) performing a second search of the query in an unstructured data source;
- g) integrating the results of the first and second searches and selecting an answer to the natural language question; and
- h) displaying the selected answer to the natural language question.

6. The method of digital data gathering for providing an answer to a natural language question, according to Claim 5, further comprising: providing a most likely answer to the natural language question.

7. The method of digital data gathering for providing an answer to a natural language question, according to Claim 5, further comprising: eliminating redundant search results and ranking search results in order of relevance.

8. The method of digital data gathering for providing an answer to a natural language question, according to Claim 5, further comprising: routing the query and identified structured data source to a structured data source manager.

9. The method of digital data gathering for providing an answer to a natural language question, according to Claim 5, further comprising: providing the structured data source in a physical data warehouse.

10. The method of digital data gathering for providing an answer to a natural language question, according to Claim 9, further comprising: identifying the structured data source via a meta-data source for the physical data warehouse.

11. The method of digital data gathering for providing an answer to a natural language question, according to Claim 5, further comprising: eliminating irrelevant words of the natural language question from use in the query.

12. The method of digital data gathering for providing an answer to a natural language question, according to Claim 5, further comprising: routing the query to an unstructured data source manager.

13. The method of digital data gathering for providing an answer to a natural language question, according to Claim 5, further comprising: displaying data related to the answer.

14. The method of digital data gathering for providing an answer to a natural language question, according to Claim 5, further comprising: accumulating search results for a specified time or specified number of results before displaying the answer.

15. The method of digital data gathering for providing an answer to a natural language question, according to Claim 5, further comprising: accumulating additional search results after displaying the answer.

16. The method of digital data gathering for providing an answer to a natural language question, according to Claim 15, further comprising: updating the ranking of the search results by incorporating the additional search results.

17. The method of digital data gathering for providing an answer to a natural language question, according to Claim 16, further comprising: providing a second display updating the ranking of the search results by incorporating the additional search results.

18. The method of digital data gathering for providing an answer to a natural language question, according to Claim 17, wherein: the second display updating the ranking of the search results is manually actuated.

19. An intranet mediator for providing a most likely answer to a natural language question, comprising:

a) a user interface with:

i) a natural language question input module for accepting natural language questions; and

ii) an answer module for display of the most likely answer;

b) a parser module for identifying the relevant concepts of the natural language question, assembling the relevant concepts of the natural language question into a query and eliminating irrelevant words of the natural language question from use in the query;

c) an unstructured data source manager for managing query input to, and accepting results from, unstructured data sources;

d) a data source selection module for accepting the query from the parser and for identifying a data source likely to contain an answer to the query; the data source selection module being connectable to a meta-data source for a physical data warehouse,

YOGESH SHARMA

- e) a dispatcher module for accepting the query from the parser and for accepting the identified data source from the data source selection module and routing the query and identified data source to a structured data source manager or an unstructured data source manager, or both;
- f) a structured data source manager for accepting the query from the dispatcher and performing a search of the query in the data warehouse and forwarding the results of the search to a results manager module;
- g) the unstructured data source manager further accepting the query and any identified unstructured data sources from the dispatcher and performing a search of the query in the identified unstructured data sources and forwarding the results of the search to a results manager; and
- h) a results manager module for accepting the results of the structured and unstructured data source searches and integrating the results of the searches and selecting the most likely answer and forwarding the most likely answer to the answer module.

20. An intranet mediator for providing a most likely answer to a natural language question according to Claim 19, further comprising: the natural language question input module being constructed and arranged for allowing the user to manually select data sources if desired.

21. An intranet mediator for providing a most likely answer to a natural language question according to Claim 19, further comprising: the answer module being constructed and arranged for display of the most likely answer and data associated therewith.

22. The method of digital data gathering for providing an answer to a natural language question, according to Claim 21, the results manager module further comprising: means for accumulating search results for a specified time or specified number of results before displaying the answer.

23. The method of digital data gathering for providing an answer to a natural language question, according to Claim 21, the results manager module further comprising: means for accumulating additional search results after displaying the answer.

24. The method of digital data gathering for providing an answer to a natural language question, according to Claim 23, the results manager module further comprising: means for updating the ranking of the search results by incorporating the additional search results.

25. The method of digital data gathering for providing an answer to a natural language question, according to Claim 24, the answer module further comprising: means for providing a second display updating the ranking of the search results by incorporating the additional search results.

26. The method of digital data gathering for providing an answer to a natural language question, according to Claim 25, further comprising: means for manually actuating the second display.

27. An intranet mediator for providing a most likely answer to a natural language question, comprising:

- a) a physical data warehouse containing structured data sources;
- b) unstructured data sources;
- c) a meta-data repository having meta-data for the structured data sources;
- d) a natural language question input module for accepting natural language queries and allowing the user to manually select data sources if desired;
- e) a parser module for identifying the relevant concepts of the natural language question, assembling the relevant concepts of the natural language question into primary query tokens and eliminating irrelevant words of the natural language question from use as primary query tokens, and for accepting results from a query expander module;

TECHNICAL SPECIFICATIONS

- f) a query expander module for accepting the primary query, determining analogous terms to the primary query tokens, and forwarding the primary query tokens and the analogous terms to an unstructured data source manager, and assembling enhanced query tokens from the results;
- g) an unstructured data source manager for managing enhanced query token input to, and accepting search results from, the unstructured data sources;
- h) a data source selection module for accepting the enhanced query from the parser module and connectable to the meta-data source for the physical data warehouse, and for identifying a data source likely to contain an answer to each of the enhanced query tokens;
- i) a dispatcher module for accepting the enhanced query tokens from the parser and for accepting the identified data sources from the data source selection module and routing the enhanced query tokens and identified data sources to a structured data source manager and an unstructured data source manager;
- j) a structured source manager for accepting the enhanced query tokens and the identified structured data sources from the dispatcher and performing a search of the enhanced query tokens in the identified structured sources and forwarding the results of the search to a results manager module;

k) the unstructured source manager further accepting the enhanced query tokens and identified unstructured data sources from the dispatcher and performing a search of the enhanced query tokens in the identified unstructured data sources and forwarding the results of the search to a results manager;

l) a results manager module for accepting the results of the structured and unstructured data source searches for each enhanced query token and integrating the results of the searches and selecting the most likely answer to the natural language question and forwarding the most likely answer to the answer module; and

m) an answer module for display of the most likely answer and associated data links.

28. The intranet mediator for providing a most likely answer to a natural language question, according to Claim 27, further comprising: the meta-data repository having meta-data for at least some of the unstructured data sources.